

**INTERNATIONAL JOURNAL OF LAW
MANAGEMENT & HUMANITIES**
[ISSN 2581-5369]

Volume 9 | Issue 5

2026

© 2026 *International Journal of Law Management & Humanities*

Follow this and additional works at: <https://www.ijlmh.com/>

Under the aegis of VidhiAagaz – Inking Your Brain (<https://www.vidhiaagaz.com/>)

This article is brought to you for free and open access by the International Journal of Law Management & Humanities at VidhiAagaz. It has been accepted for inclusion in the International Journal of Law Management & Humanities after due review.

In case of **any suggestions or complaints**, kindly contact support@vidhiaagaz.com.

To submit your Manuscript for Publication in the **International Journal of Law Management & Humanities**, kindly email your Manuscript to submission@ijlmh.com.

Accountability and Transparency in AI Systems: A Human Rights Perspective

NIKITA SHARMA* AND DR. SANDEEP SAINI**

ABSTRACT

The proliferation of artificial intelligence across public and private sectors has positioned algorithmic decision-making at the heart of consequential outcomes, from criminal sentencing and credit scoring to healthcare allocation and employment screening. Yet the opacity of AI systems, combined with fragmented governance structures, has created a significant accountability deficit that threatens the realisation of fundamental human rights. This paper addresses the central research question: to what extent do existing human rights legal frameworks impose enforceable accountability and transparency obligations on states and corporate actors deploying AI systems? Using a doctrinal legal analysis methodology, the paper systematically examines primary legal instruments, including the UN Guiding Principles on Business and Human Rights, the International Covenant on Civil and Political Rights, the EU AI Act, and UNESCO's Recommendation on the Ethics of AI, to assess their capacity to govern AI-driven decision-making in rights-sensitive contexts. With special reference to the Indian constitutional framework, anchored in the Supreme Court's landmark recognition of privacy as a fundamental right in Justice K.S. Puttaswamy (Retd.) v. Union of India (2017), and to the emerging architecture of the Digital Personal Data Protection Act, 2023, the paper situates India's governance challenges within the global accountability debate. The paper identifies critical governance gaps, including the absence of enforceable audit mechanisms, inadequate redress pathways for rights-affected individuals, and the difficulty of attributing legal responsibility across complex AI supply chains. It concludes by proposing a human rights-by-design governance model that embeds accountability and transparency obligations across the full AI lifecycle, offering a principled framework for reconciling technological innovation with fundamental rights protection.

Keywords: artificial intelligence, accountability, transparency, human rights, AI governance, EU AI Act, algorithmic decision-making, doctrinal legal analysis, UN Guiding Principles, Puttaswamy, DPDPA, India

* Author is a Ph.D. Scholar at Geeta Global School of Law, Geeta University, Panipat, Haryana, India.

** Author is an Associate Professor and Head at Geeta Global School of Law, Geeta University, Panipat, Haryana, India.

I. INTRODUCTION

Artificial intelligence has become an infrastructure of power. Deployed across judicial, administrative, financial, and social welfare systems, AI systems determine who receives bail, who is approved for credit, which patients are prioritised for treatment, and which job applicants advance to interview.¹ The consequences of these determinations are profound and, often, irreversible. Yet the processes by which AI systems reach their conclusions remain, in most jurisdictions and in most deployment contexts, opaque, unaudited, and practically unreviewable.²

This accountability deficit is not merely a technical failure. It is a human rights failure. The rights to equality, due process, privacy, and dignity, recognised in binding international instruments and, in India, guaranteed as fundamental rights under Part III of the Constitution, are structurally compromised when decisions affecting their exercise are delegated to algorithmic systems that cannot be interrogated, understood, or effectively challenged. The United Nations Special Rapporteur on Freedom of Opinion and Expression has warned that algorithmic systems may be so complex as to frustrate explanation, and that where an AI system is used by a public agency, a vendor's refusal to be transparent about its operation is incompatible with that agency's own accountability obligations.³

India presents a particularly critical case study. The National Strategy for Artificial Intelligence, published by NITI Aayog in 2018, positions India as an aspirant AI superpower seeking to leverage AI for inclusive growth across agriculture, healthcare, smart cities, and financial services.⁴ NITI Aayog has since published principles for responsible AI.⁵ Yet as AI systems penetrate the operations of welfare administration, policing, judicial support, and financial services in India, the constitutional framework, particularly the proportionality standard articulated in *Puttaswamy*, demands that these deployments be subject to enforceable accountability and transparency obligations that current law does not provide.⁶

¹ VIRGINIA EUBANKS, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR* 8–12 (St. Martin's Press 2018).

² KATE CRAWFORD, *ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE* 211–14 (Yale Univ. Press 2021).

³ Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, *Promotion and Protection of the Right to Freedom of Opinion and Expression*, ¶¶ 6, 55, transmitted by Note of the Secretary-General, U.N. Doc. A/73/348 (Aug. 29, 2018) [hereinafter *SR Freedom of Expression Report*].

⁴ NITI AAYOG, *National Strategy for Artificial Intelligence* 3–7 (Gov't of India 2018), <https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf> [hereinafter *NITI Aayog AI Strategy*].

⁵ NITI AAYOG, *Responsible AI #AIForAll: Approach Document for India, Part 1: Principles for Responsible AI* 41–42 (Gov't of India 2021), <https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf> [hereinafter *NITI Aayog Responsible AI Part 1*].

⁶ *Justice K.S. Puttaswamy (Retd.) v. Union of India*, (2017) 10 SCC 1 (India) [hereinafter *Puttaswamy*].

Following this introduction, the paper proceeds in six parts. Part II surveys the human rights foundations of AI accountability obligations. Part III examines the principal international and regional governance instruments. Part IV analyses the Indian constitutional and statutory framework. Part V identifies the central governance gaps: audit deficits, redress failures, and supply chain accountability problems. Part VI proposes a human rights-by-design lifecycle governance model. Part VII concludes.

II. HUMAN RIGHTS FOUNDATIONS OF AI ACCOUNTABILITY

A. The International Covenant on Civil and Political Rights

The International Covenant on Civil and Political Rights (ICCPR) constitutes the principal binding framework of civil and political rights obligations applicable to state conduct in the domain of AI governance. Its Articles 2(3), 14, and 17, establishing the rights to an effective remedy, a fair trial, and privacy respectively, collectively generate a set of minimum requirements that state-operated or state-sanctioned AI decision-making systems must satisfy.⁷

The Human Rights Committee's General Comment No. 31 makes clear that the obligation to provide effective remedies under Article 2(3) is not limited to violations caused by direct state action but extends to situations in which states fail to exercise due diligence to prevent violations by private actors.⁸ Where private corporations deploy AI systems with foreseeable rights impacts in employment, credit allocation, or insurance, and states fail to regulate or audit such deployments, the state may incur responsibility for the resulting rights violations. The Committee's General Comment No. 16 on privacy further establishes that the holding of personal data in automatic data files generates specific obligations of transparency, and that individuals must be able to ascertain what personal data is held about them and for what purposes.⁹

The right to a fair hearing under Article 14, which the Committee has interpreted to encompass not only judicial proceedings but also certain administrative determinations, such as those concerning social security benefits, imposes requirements of reasoning and reviewability that are incompatible with opaque algorithmic decision-making in contexts such as benefit denial,

⁷ International Covenant on Civil and Political Rights arts. 2(3), 14, 17, Dec. 16, 1966, 999 U.N.T.S. 171 [hereinafter ICCPR].

⁸ U.N. HUMAN RIGHTS COMMITTEE, *General Comment No. 31: The Nature of the General Legal Obligation Imposed on States Parties to the Covenant*, ¶ 8, U.N. Doc. CCPR/C/21/Rev.1/Add.13 (May 26, 2004).

⁹ ICCPR, *supra* note 7, art. 17; U.N. HUMAN RIGHTS COMMITTEE, *General Comment No. 16: The Right to Respect of Privacy, Family, Home and Correspondence, and Protection of Honour and Reputation*, ¶ 10 (Apr. 8, 1988), U.N. Doc. HRI/GEN/1/Rev.1 (1994).

licence revocation, or regulatory sanction.¹⁰ A system that denies a person a welfare benefit on the basis of an algorithmic risk score, without explaining the score or providing a meaningful opportunity to challenge it, *prima facie* violates Article 14 obligations.

B. The UN Guiding Principles on Business and Human Rights

The UN Guiding Principles on Business and Human Rights (UNGPs), endorsed by the Human Rights Council in 2011, establish the foundational framework for corporate accountability in the human rights domain. Their three-pillar architecture (the state duty to protect, the corporate responsibility to respect, and access to remedy) maps directly onto the principal accountability challenges of AI governance.¹¹

The corporate responsibility to respect human rights under Pillar II is operationalised through human rights due diligence: a continuous process of identifying, preventing, mitigating, and accounting for actual and potential adverse human rights impacts across the full value chain of corporate activities.¹² Applied to AI systems, due diligence requires corporations to assess the rights impacts of algorithmic systems before deployment, monitor those impacts during operation, and remediate harms as they arise. The UNGPs' emphasis on transparency, requiring corporations to communicate externally about how they address their human rights impacts, provides normative grounding for disclosure and audit obligations in AI governance.

Pillar III's access to remedy framework requires both judicial and non-judicial mechanisms to be available, accessible, and effective for individuals whose rights are adversely affected by corporate conduct.¹³ The application of this framework to AI-driven decisions raises specific challenges: the affected individual may not know that an algorithmic system was involved in the determination that harmed them; may lack the technical literacy to understand or contest its operation; and may face structural barriers (cost, complexity, information asymmetry) that render formal grievance mechanisms practically ineffective. These challenges are acutely present in the Indian context, where affected populations frequently include marginalised communities with limited access to legal resources.

¹⁰ U.N. HUMAN RIGHTS COMMITTEE, *General Comment No. 32: Article 14: Right to Equality Before Courts and Tribunals and to a Fair Trial*, ¶¶ 15–16, U.N. Doc. CCPR/C/GC/32 (Aug. 23, 2007).

¹¹ *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*, Principles 11–15, U.N. Doc. A/HRC/17/31, annex (Mar. 21, 2011), endorsed by H.R.C. Res. 17/4 (June 16, 2011), https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinessshr_en.pdf [hereinafter UNGPs].

¹² UNGPs, *supra* note 11, Principle 17 (human rights due diligence encompasses identification, prevention, mitigation, and accounting for impacts).

¹³ UNGPs, *supra* note 11, Principles 25–31 (access to remedy pillar).

C. UNESCO's Recommendation on the Ethics of AI

UNESCO's Recommendation on the Ethics of Artificial Intelligence, adopted unanimously by the General Conference in November 2021, constitutes the first global normative instrument on AI ethics and provides the most comprehensive articulation of transparency and accountability as human rights-grounded governance principles.¹⁴

The Recommendation establishes explainability, the capacity of an AI system's outcomes to be made intelligible to affected persons, as a core component of the transparency principle, and distinguishes between transparency about the AI system as such (its design, training data, and operational parameters) and transparency about specific decisions (the reasons for a particular outcome affecting a specific individual).¹⁵ Both dimensions are treated as rights-grounded obligations, not merely aspirational technical standards. The Recommendation further calls for impact assessments, independent oversight mechanisms, and the right to human review of AI decisions affecting fundamental interests, all elements that the paper's proposed governance model incorporates.

III. REGIONAL AND INTERNATIONAL GOVERNANCE INSTRUMENTS

A. The EU AI Act

The EU Artificial Intelligence Act (Regulation (EU) 2024/1689), which entered into force on 1 August 2024, represents the world's first comprehensive binding regulatory framework for AI systems and sets a global benchmark against which other governance regimes, including India's emerging framework, must be measured.¹⁶ Its obligations apply in stages: the prohibited practices from 2 February 2025, the rules on general-purpose AI models from 2 August 2025 and, following the postponement enacted by the Digital Omnibus on AI in July 2026, the high-risk obligations from 2 December 2027 for the stand-alone systems listed in Annex III and from 2 August 2028 for systems embedded in products covered by Annex I.¹⁷

¹⁴ UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, ¶¶ 4, 42–47, U.N. Doc. SHS/BIO/PI/2021/1 (Nov. 23, 2021), <https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence> [hereinafter UNESCO AI Recommendation].

¹⁵ UNESCO AI Recommendation, *supra* note 14, ¶¶ 37–41 (transparency and explainability; ¶ 40 defining explainability).

¹⁶ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), arts. 9–17, 2024 O.J. (L 1689) 1, <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> [hereinafter EU AI Act].

¹⁷ EU AI Act, *supra* note 16, art. 113; Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 Amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as Regards the Simplification of the Implementation of Harmonised Rules on Artificial Intelligence (Digital Omnibus on AI), 2026 O.J. (L 1744) 1, <https://eur-lex.europa.eu/eli/reg/2026/1744/oj>.

The Act's risk-based architecture classifies AI systems into four tiers (unacceptable risk, which is prohibited; high risk; limited risk; and minimal risk), with the most stringent obligations falling on high-risk systems, defined in part by reference to their deployment in the Annex III contexts, which include critical infrastructure, employment, education, essential services, law enforcement, migration, and the administration of justice.¹⁸ High-risk systems are subject to mandatory requirements of risk management, data governance, technical documentation, transparency and provision of information to deployers, human oversight, accuracy, robustness, and cybersecurity: a comprehensive lifecycle accountability framework.

Of particular significance for the human rights analysis are Articles 13 and 14, which impose transparency and human oversight obligations on high-risk AI systems.¹⁹ Article 13 requires that high-risk AI systems be designed to enable deployers to interpret their outputs and use them appropriately, including through the provision of instructions for use detailing system purpose, performance characteristics, and known limitations. Article 14 mandates that high-risk systems be designed to allow effective oversight by natural persons during deployment, including the ability to decide not to use the system's output or to override it. These provisions translate the abstract transparency principle into concrete technical and operational requirements.

The Act's risk management obligations under Article 9 require providers to establish and maintain a risk management system throughout the lifecycle of the AI system, encompassing identification and analysis of known and foreseeable risks, evaluation of risks that cannot be eliminated, and adoption of appropriate risk management measures.²⁰ This lifecycle orientation, recognising that AI risks are dynamic and evolve with deployment context, user adaptation, and model updates, is a significant advance over static, point-in-time compliance models and directly informs the governance framework proposed in Part VI.

B. The GDPR and the contested right to explanation

The General Data Protection Regulation (GDPR) provides an important, if imperfect, foundation for transparency rights in AI-driven decision-making through its Articles 13, 14, 15, and 22.²¹ Article 22 grants data subjects a right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning them or

¹⁸ EU AI Act, *supra* note 16, arts. 6–7 & annex III (classification of high-risk AI systems and Annex III categories including critical infrastructure, employment, and access to essential services).

¹⁹ EU AI Act, *supra* note 16, arts. 13–14 (transparency and human oversight requirements for high-risk AI systems).

²⁰ EU AI Act, *supra* note 16, art. 9 (risk management system requirements throughout AI lifecycle).

²¹ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation), arts. 13–15, 22, 2016 O.J. (L 119) 1, <https://eur-lex.europa.eu/eli/reg/2016/679/oj> [hereinafter GDPR].

similarly significantly affects them, unless the decision is necessary for a contract, authorised by Union or Member State law, or based on explicit consent.²² Where such a decision rests on contract or explicit consent, the data subject is entitled at least to obtain human intervention, to express his or her point of view, and to contest the decision; where it rests on law, that law must itself lay down suitable safeguards.

The GDPR's transparency obligations under Articles 13, 14, and 15, requiring meaningful information about the logic involved in automated decision-making and the envisaged significance and consequences for the data subject, have been criticised for their limited enforceability in the face of technically complex AI systems. Recital 71 refers to a right "to obtain an explanation of the decision reached", but recitals are not binding and the operative text contains no such express right. Scholars including Wachter, Mittelstadt, and Floridi have argued on this basis that no genuine right to explanation exists in the GDPR's current text, but only a right to generic information about the logic of processing.²³ That reading remains contested, and the Court of Justice has since held that Article 15(1)(h) entitles the data subject to an explanation, in concise and intelligible form, of the procedure and principles actually applied to his or her personal data.²⁴ The uncertainty that remains reinforces the case for the more prescriptive algorithmic transparency requirements of the EU AI Act, which supplement the GDPR's data protection framework with purpose-specific accountability obligations.

IV. THE INDIAN CONSTITUTIONAL AND STATUTORY FRAMEWORK

A. The *Puttaswamy* foundation

The Supreme Court of India's nine-judge bench decision in *Justice K.S. Puttaswamy (Retd.) v. Union of India* (2017) constitutes the jurisprudential bedrock upon which any accountability framework for AI systems in India must be constructed.²⁵ The unanimous recognition of privacy as a fundamental right under Article 21 of the Constitution was not merely declaratory: it generated a positive constitutional obligation on the state to create conditions, including a legislative framework, that protect individual privacy against technological intrusions.

The proportionality standard articulated in *Puttaswamy*, which requires that any limitation of the right to privacy satisfy the three-fold requirement of legality, a legitimate state aim, and

²² GDPR, *supra* note 21, art. 22(1)–(3) & recital 71 (rights in relation to automated individual decision-making, including profiling).

²³ Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIVACY L. 76, 77–80 (2017), <https://doi.org/10.1093/idpl/ix005>.

²⁴ Case C-203/22, *CK v. Magistrat der Stadt Wien (Dun & Bradstreet Austria)*, ECLI:EU:C:2025:117 (Feb. 27, 2025).

²⁵ *Puttaswamy*, *supra* note 6.

proportionality ensuring a rational nexus between the objects and the means adopted, provides a powerful constitutional tool for challenging AI-driven administrative decisions that affect individuals without adequate justification, transparency, or opportunity for redress.²⁶ Applied to AI systems deployed in welfare administration, policing, or financial services, the proportionality analysis requires that the state demonstrate, for each AI deployment, that the intrusion into privacy (and associated rights to dignity and equality) is not greater than strictly necessary for the legitimate aim pursued. The absence of explainability mechanisms, audit obligations, or redress pathways in current AI deployments in India would, on this analysis, render many such systems constitutionally vulnerable.

The recognition in *Puttaswamy* of informational privacy, including the individual's right to control the dissemination of personal information, as a component of the fundamental right to privacy further grounds a constitutional obligation of AI transparency: where algorithmic systems make decisions about individuals based on personal data, those individuals have a constitutional entitlement to understand the basis of those decisions.²⁷

B. The Digital Personal Data Protection Act, 2023

The Digital Personal Data Protection Act, 2023 (DPDPA), enacted in August 2023, constitutes India's first comprehensive data protection legislation and the primary statutory instrument that will govern the processing of personal data by AI systems.²⁸ Its commencement is staggered: the Digital Personal Data Protection Rules, 2025 were notified on 13 November 2025, when the provisions establishing the Data Protection Board came into force, but the core obligations and rights in Sections 3 to 17 take effect only eighteen months later, in May 2027.²⁹ Its foundational provisions on consent, purpose limitation, and data minimisation broadly track GDPR-influenced international standards and, once in force, will provide a partial basis for challenging AI-driven data processing that fails to respect individual rights.

However, the DPDPA's adequacy as an AI accountability instrument is compromised by two structural features. First, Section 17(2)(a) empowers the Central Government to exempt, by

²⁶ *Puttaswamy*, *supra* note 6, ¶¶ 310, 325 (Chandrachud J., for the plurality) (three-fold requirement of legality, need defined in terms of a legitimate state aim, and proportionality ensuring a rational nexus between the objects and the means adopted).

²⁷ *Puttaswamy*, *supra* note 6 (Kaul J.) (the right of an individual to control the dissemination of personal information as a component of the fundamental right to privacy, and the need for a data protection framework).

²⁸ Digital Personal Data Protection Act, No. 22 of 2023, §§ 4–6, 12–13 (India) [hereinafter DPDPA].

²⁹ MINISTRY OF ELECTRONICS AND INFORMATION TECHNOLOGY, Notification Appointing the Dates of Commencement of the Digital Personal Data Protection Act, 2023, Gazette of India, Extraordinary, pt. II sec. 3(i) (Nov. 13, 2025) (India) (bringing §§ 18–26 into force on Nov. 13, 2025 and §§ 3–17 into force eighteen months thereafter); The Digital Personal Data Protection Rules, 2025, Gazette of India, Extraordinary, pt. II sec. 3(i) (Nov. 13, 2025) (India).

notification, any instrumentality of the State from the Act's provisions in the interests of the sovereignty and integrity of India, the security of the State, public order, and other listed grounds.³⁰ The breadth of these exemptions, which could extend to government-operated AI systems in welfare administration, policing, and border control, risks creating a two-tier accountability regime in which private sector AI is subject to data protection obligations while government AI operates in a regulatory vacuum. This is particularly concerning because many of the most rights-sensitive AI deployments, such as those in welfare administration and policing, are undertaken by state actors.

Second, the DPDPA does not contain provisions specifically addressing automated decision-making or algorithmic accountability. Unlike GDPR Article 22, it creates no right against solely automated decisions, no right to human review, and no obligation of algorithmic explanation. The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 address the use of automated tools by significant social media intermediaries, subject to human oversight, but impose no accountability requirements on AI decision-making systems in other sectors.³¹ The result is a statutory landscape that, notwithstanding the constitutional foundation of *Puttaswamy*, leaves AI accountability obligations substantially underspecified.

NITI Aayog has published principles for responsible AI emphasising safety, equality, inclusivity, privacy, transparency, and accountability, and MeitY, after a public consultation, issued the India AI Governance Guidelines in November 2025.³² These soft-law instruments represent valuable normative development but do not substitute for binding legislative obligations. The gap between India's AI ambition (on an Accenture estimate relied on in NITI Aayog's National Strategy, AI could add 1.3 percentage points to India's annual growth rate by 2035)³³ and the accountability architecture required to make that deployment rights-compatible is a defining challenge for Indian law in the coming legislative cycle.

³⁰ DPDPA, *supra* note 28, § 17(2)(a) (power of the Central Government to exempt, by notification, instrumentalities of the State in the interests of sovereignty, security of the State, public order and other listed grounds, potentially broad enough to insulate government AI systems from accountability requirements).

³¹ The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, r. 4(4) (India) (requiring significant social media intermediaries to endeavour to deploy automated tools, subject to appropriate human oversight and periodic review, but not imposing substantive algorithmic accountability on AI decision-making systems generally).

³² NITI Aayog Responsible AI Part 1, *supra* note 5, at 41–42; *see also* MINISTRY OF ELECTRONICS AND INFORMATION TECHNOLOGY, *Report on AI Governance Guidelines Development* (Jan. 6, 2025) (released for public consultation); MINISTRY OF ELECTRONICS AND INFORMATION TECHNOLOGY, *India AI Governance Guidelines* (Nov. 5, 2025).

³³ NITI Aayog AI Strategy, *supra* note 4, at 18 (citing Accenture's estimate that AI could boost India's annual growth rate by 1.3 percentage points by 2035).

V. CRITICAL GOVERNANCE GAPS

A. The audit deficit

The most structurally significant gap in existing AI governance frameworks is the absence of mandatory, independent, *ex ante* and *ex post* algorithmic auditing. The “black box” problem, the inability of affected persons, regulators, and courts to understand how an AI system produces its outputs, is not merely a technical inconvenience but a fundamental obstacle to accountability.³⁴ Without audit mechanisms that assess system accuracy, bias, and inferential scope before deployment and monitor these parameters continuously during operation, the transparency obligations of the ICCPR, the UNGPs, and the EU AI Act remain aspirational rather than enforceable.

The challenge of algorithmic transparency takes two forms that governance frameworks must address separately. The first is process transparency: disclosure of what data the system was trained on, what optimisation objective it pursues, and what its documented performance characteristics are. The second is outcome transparency: the ability to explain, in terms comprehensible to the affected individual, why a specific decision was reached in their case.³⁵ These are distinct obligations. A system may be process-transparent, its training data and architecture fully documented, yet outcome-opaque, generating predictions that no post-hoc explanation technique can reliably attribute to specific input features.

The GDPR’s transparency provisions have been criticised for conflating these two dimensions and failing to secure genuine outcome explainability.³⁶ The EU AI Act’s Article 13 advances the analysis by requiring that high-risk AI systems be designed for interpretability by deployers, but does not itself guarantee that affected individuals can access case-specific explanations, a gap the Act addresses only through the narrower right to explanation in Article 86.³⁷ In the Indian context, neither the DPDPA nor any sector-specific regulatory instrument imposes algorithmic audit obligations, and India’s Data Protection Board, the statutory body tasked with enforcement, although established in law in November 2025, was still without a Chairperson or Members in May 2026,³⁸ leaving the entire accountability architecture operationally inert.

³⁴ FRANK PASQUALE, *THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION* 3–8 (Harvard Univ. Press 2015).

³⁵ Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1508–12.

³⁶ GDPR, *supra* note 21, arts. 13–14 (information to be provided; no guarantee of technical interpretability of model logic).

³⁷ EU AI Act, *supra* note 16, art. 13(1) (high-risk AI systems shall be designed and developed in such a way as to ensure that their operation is sufficiently transparent to enable deployers to interpret a system’s output and use it appropriately); *id.* art. 86.

³⁸ MINISTRY OF ELECTRONICS AND INFORMATION TECHNOLOGY, *Appointment to the Post of Chairperson and Other Members in the Data Protection Board of India*, F. No. 2(1)/2026-Pers.I (May 6, 2026) (India),

B. The redress deficit

The second critical gap is the inadequacy of redress pathways for individuals whose rights are adversely affected by AI-driven decisions. Effective remedy, guaranteed under ICCPR Article 2(3) and grounded in the *Puttaswamy* proportionality framework, requires not merely formal access to a complaints body but a practically effective mechanism: one that is accessible to affected individuals, capable of providing timely relief, and empowered to impose meaningful consequences on rights-violating actors.³⁹

Existing frameworks fall significantly short of this standard. The EU AI Act provides rights to lodge a complaint and to obtain an explanation under Articles 85 and 86, but enforcement is entrusted to national market surveillance authorities whose capacity, prioritisation, and accessibility vary significantly across member states.⁴⁰ The DPDPA establishes a right of grievance redressal with the Data Fiduciary (the entity that determines the purpose and means of processing) and ultimate recourse to the Data Protection Board, but Section 13 does not come into force until May 2027 and the Board had yet to be staffed in mid-2026, so this pathway is not yet operative.⁴¹ For rural and marginalised communities in India, those most frequently subject to automated welfare and policing decisions and least able to navigate formal complaints processes, the practical inaccessibility of formal redress mechanisms is compounded by digital illiteracy, language barriers, and structural power asymmetries.

A broader challenge of algorithmic redress in the Global South is that governance frameworks developed primarily in high-income, high-digital-literacy contexts do not translate without adaptation. India's AI accountability framework must design redress mechanisms appropriate to a context of linguistic diversity, uneven digital access, and concentrated administrative power if the *Puttaswamy* promise of enforceable privacy rights is to be realised in practice.

C. Supply chain accountability

The third governance gap concerns the attribution of legal responsibility across complex AI supply chains. An AI system deployed in public welfare administration may involve a government deployer, a private technology vendor that supplied the model, a data provider, a cloud infrastructure operator, and a third-party auditor, none of whom individually controls the

<https://www.meity.gov.in/static/uploads/2026/05/cd481c027470b420b4cb85fb40a91c53.pdf>.

³⁹ ICCPR, *supra* note 7, art. 2(3) (effective remedy obligation extends to all violations of Covenant rights, including those caused by private actors where the State fails to exercise due diligence); *see* General Comment No. 31, *supra* note 8, ¶ 8.

⁴⁰ EU AI Act, *supra* note 16, arts. 85–86 (right to lodge a complaint with a market surveillance authority; right to an explanation of individual decision-making based on the output of a high-risk AI system listed in Annex III).

⁴¹ DPDPA, *supra* note 28, § 13 (right of grievance redressal with the Data Fiduciary); *see supra* notes 29, 38 (commencement of § 13 deferred to May 2027; Board without a Chairperson or Members as of May 2026).

full system whose outputs produce rights impacts.⁴² The corporate responsibility to respect human rights under the UNGPs extends across the value chain of corporate activities, but translating this principle into enforceable legal obligations in AI supply chains requires regulatory frameworks to specify which actors bear what obligations at which points in the lifecycle.

The EU AI Act's value-chain obligations under Articles 16 and 23 to 26, allocating responsibilities among providers (who develop the system and place it on the market), importers, distributors, and deployers (who use the system), represent an important advance, but leave significant gaps in contexts where the deployer is a government entity that may benefit from sovereign immunities, or where the AI system is developed and operated by a single integrated actor that combines both provider and deployer roles.⁴³ In India, where government agencies routinely procure AI systems from domestic and foreign technology companies under public procurement contracts that do not standardly incorporate human rights due diligence requirements, the supply chain accountability gap is particularly acute. The constitutional obligations generated by *Puttaswamy* bind the state directly, but the contractual and regulatory instruments through which those obligations could be transmitted downstream to private vendors do not currently exist.

VI. A HUMAN RIGHTS-BY-DESIGN LIFECYCLE GOVERNANCE MODEL

A. Foundational principles

The governance framework proposed in this paper is organised around the principle of human rights-by-design: the requirement that accountability and transparency obligations be embedded in the architecture of AI systems from the earliest design stages, rather than retrofitted as compliance obligations after deployment.⁴⁴ This principle, reflected in the UNESCO Recommendation's call for respect for human rights throughout the life cycle of AI systems and developed in the EU AI Act's lifecycle accountability architecture, is consistent with the *Puttaswamy* court's recognition that constitutional obligations of privacy protection require proactive legislative and regulatory action, not merely reactive judicial enforcement.

The framework is further grounded in the UNGPs' concept of continuous human rights due diligence: a living process that evolves as systems are updated, as deployment contexts change,

⁴² UNGPs, *supra* note 11, Principles 12–15 (corporate responsibility to respect covers all internationally recognised human rights across the full value chain of operations).

⁴³ EU AI Act, *supra* note 16, arts. 16, 23–26 (obligations of providers, importers, distributors and deployers, and responsibilities along the AI value chain).

⁴⁴ UNESCO AI Recommendation, *supra* note 14, ¶¶ 62–67 (requiring, at ¶ 65, that the actions of AI actors be consistent with international human rights law throughout the life cycle of AI systems).

and as new impacts are identified.⁴⁵ Unlike static compliance models that treat accountability as a box to be ticked at the point of deployment, the lifecycle model proposed here treats accountability as an ongoing operational obligation that persists throughout the system's existence and survives changes in ownership, purpose, or deployment context.

B. Pre-deployment: rights impact assessment

The first phase of the lifecycle model requires that all AI systems intended for deployment in rights-sensitive contexts undergo a mandatory human rights impact assessment (HRIA) prior to deployment. Drawing on the UNGPs' due diligence framework and the EU AI Act's conformity assessment model,⁴⁶ the HRIA must: identify all human rights that could foreseeably be affected by the system's operation; assess the probability, severity, and reversibility of potential impacts; document the system's training data, optimisation objectives, and known limitations; and demonstrate that less rights-intrusive alternatives have been considered and rejected for documented reasons.

In the Indian context, the HRIA obligation should be extended to cover not only the constitutional rights framework, namely Article 14 (equality), Article 19 (freedom of expression), and Article 21 (life, liberty, privacy, and dignity), but also the specific vulnerabilities of historically marginalised communities, including Scheduled Castes and Scheduled Tribes, religious minorities, women, and persons with disabilities, whose disproportionate exposure to algorithmic harm reflects and reinforces structural inequalities.⁴⁷ NITI Aayog's responsible AI principles acknowledge these concerns but stop short of mandating HRIAs as a precondition for deployment.

C. Deployment: transparency and human oversight

During deployment, the framework requires two categories of obligation. First, operational transparency: AI systems used in rights-sensitive decisions must generate, and make available to affected persons, case-specific explanations of the basis for decisions. These explanations must be expressed in terms comprehensible to non-expert individuals and must be available in the official languages of the jurisdiction, a requirement of particular significance in India's

⁴⁵ UNGPs, *supra* note 11, Principle 17(c) (human rights due diligence should be ongoing, recognising that human rights risks may change over time).

⁴⁶ EU AI Act, *supra* note 16, art. 43 (conformity assessment of high-risk AI systems; assessment involving a notified body is required only for certain biometric systems under point 1 of Annex III, the other Annex III systems following internal control).

⁴⁷ NITI AAYOG, *Responsible AI #AIForAll: Approach Document for India, Part 2: Operationalizing Principles for Responsible AI* 14–18 (Gov't of India 2021), <https://www.niti.gov.in/sites/default/files/2021-08/Part2-Responsible-AI-12082021.pdf> [hereinafter NITI Aayog Responsible AI Part 2].

multilingual context, where a national AI accountability framework that operates only in English effectively denies meaningful transparency to the majority of AI-affected citizens.⁴⁸

Second, human oversight: all high-stakes AI decisions, including decisions on welfare entitlements, creditworthiness, bail or parole, and employment, must be subject to mandatory human review before being given effect, with the reviewing official required to certify that they have considered the AI system's output but have independently assessed its applicability to the specific case.⁴⁹ This requirement, operationalising the EU AI Act's Article 14 human oversight standard, is consistent with the due process dimension of Article 14 of the ICCPR and the *Puttaswamy* proportionality standard's requirement that intrusions on fundamental rights be demonstrably necessary rather than algorithmically convenient.

D. Post-deployment: audit, redress, and decommissioning

The post-deployment phase imposes three sets of obligations. First, continuous algorithmic auditing: AI systems in high-risk deployment contexts must be subject to periodic independent audits that assess accuracy, bias, inferential scope, and rights compliance, with audit reports published in accessible formats and submitted to relevant regulatory authorities. The absence of an operational Data Protection Board, and of any AI audit competence in the Board's statutory mandate, is a critical institutional lacuna that legislative action must urgently address.⁵⁰

Second, effective redress: regulators must establish grievance mechanisms that are accessible to rights-affected individuals without legal representation, that operate within defined timeframes, and that are empowered to order cessation of processing, rectification of decisions, and compensation for rights violations. The UNGPs' effectiveness criteria for non-judicial grievance mechanisms (legitimacy, accessibility, predictability, equitability, rights-compatibility, transparency, and continuous learning) provide an operational benchmark.⁵¹

Third, decommissioning accountability: when AI systems are discontinued, the governance obligations do not terminate. Data deletion obligations, model documentation retention requirements, and continuing rights to explanation for decisions made by the decommissioned

⁴⁸ UNESCO AI Recommendation, *supra* note 14, ¶ 38 (individuals should be able to access the reasons for a decision affecting their rights).

⁴⁹ EU AI Act, *supra* note 16, art. 14(4)(d) (natural persons assigned human oversight must be enabled to decide not to use the system or to disregard, override or reverse its output).

⁵⁰ DPDPA, *supra* note 28, §§ 18, 27 (establishing the Board and defining its powers by reference to personal data breaches and non-compliance with the Act, with no algorithmic audit function); *see supra* note 38.

⁵¹ UNGPs, *supra* note 11, Principle 31 (effectiveness criteria for non-judicial grievance mechanisms).

system must be maintained for defined periods.⁵² In the surveillance capitalism paradigm identified by Zuboff, the extractive value of AI systems persists beyond their operational life through the data assets they have generated and the behavioural predictions they have encoded.⁵³ A decommissioning accountability framework ensures that these residual risks are governed, not abandoned.

E. India-specific implementation imperatives

The implementation of the proposed framework in India requires a constellation of regulatory and legislative actions. First, the full operationalisation of the Data Protection Board under the DPDPA, with a mandate extended to encompass algorithmic accountability and AI audit functions. Second, the enactment of a sector-specific AI Governance and Accountability Act that imposes HRIA obligations, audit requirements, and human oversight standards on AI deployments in government and high-risk private contexts, building on the NITI Aayog responsible AI principles but translating them into binding obligations.⁵⁴

Third, the development of constitutional litigation strategies that deploy the *Puttaswamy* proportionality test to challenge specific AI deployments that fail to satisfy the requirements of legality, legitimate aim, and proportionality. The constitutional courts of India, both the Supreme Court and the High Courts, have demonstrated a willingness to scrutinise technology-mediated administrative action, as the Aadhaar litigation, including challenges concerning benefit exclusions and biometric data, illustrates.⁵⁵ These judicial interventions, while important, are reactive and piecemeal. A proactive legislative framework is essential to ensure that the constitutional right to privacy recognised in *Puttaswamy* translates into systematic accountability for AI-driven decisions.

VII. CONCLUSION

Accountability and transparency in AI systems are not merely desirable governance objectives. They are, this paper has argued, binding obligations of international human rights law: obligations that flow from the ICCPR's guarantees of effective remedy, fair hearing, and privacy; from the UNGPs' framework of corporate human rights due diligence; from UNESCO's global normative commitment to rights-grounded AI ethics; and, in India, from the constitutional imperative of *Puttaswamy*'s proportionality framework.

⁵² Eubanks, *supra* note 1, at 211–15 (decommissioning accountability: the rights harms of legacy AI systems that continue to affect individuals after formal discontinuation).

⁵³ SHOSHANA ZUBOFF, *THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER* 502–07 (PublicAffairs 2019).

⁵⁴ NITI Aayog Responsible AI Part 2, *supra* note 47, at 30–34.

⁵⁵ *Justice K.S. Puttaswamy (Retd.) v. Union of India*, (2019) 1 SCC 1 (India) (the Aadhaar judgment).

The existing governance landscape, dominated by the EU AI Act's sophisticated but geographically limited framework, the GDPR's data protection obligations, and India's emerging but incomplete DPDPA architecture, leaves substantial gaps. The audit deficit, the redress deficit, and the supply chain accountability gap collectively constitute a governance failure that exposes the most vulnerable individuals and communities to uncontrolled algorithmic power: automated decisions that shape life chances without explanation, challenge, or meaningful human oversight.

The human rights-by-design lifecycle governance model proposed in this paper offers a principled response to this failure. By embedding accountability and transparency obligations across the full AI lifecycle, from design and rights impact assessment through deployment, audit, and decommissioning, and by grounding those obligations in the constitutional framework of *Puttaswamy* and the normative architecture of international human rights law, the model provides a framework within which India can aspire to be not only an AI superpower but an AI accountability leader. The two ambitions are not in conflict. They are inseparable.

The convergence of digital governance and human rights that defines the contemporary regulatory moment demands frameworks that are technically literate, institutionally robust, and constitutionally grounded. Smart grids, predictive policing, algorithmic welfare administration, and AI-assisted judicial decision-making are not abstract future scenarios in India; they are present realities. The window for embedding rights-protective accountability mechanisms into the architecture of these systems is not unlimited. Legislative action, regulatory capacity-building, and constitutional litigation must proceed with urgency and in coordination if the promise of *Puttaswamy*, that the right to privacy is a fundamental right inseparable from human dignity, is to be honoured in the age of artificial intelligence.
